

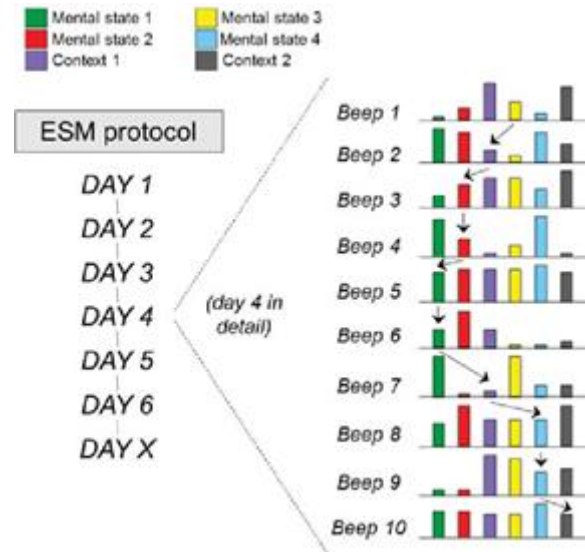
Knowledge graph representation for experience sampling data: A proof of concept

Ammar Ammar
Lianne Ippel
Remzi Celebi
Philippe Delespaul

Making life easier: Demo session 1
20-11-2020



Experience-sampling Methods (ESM)

- A form of moment-to-moment data collection.
- Collecting self-report responses from people in natural settings.
- Measure participants' feelings, thoughts, actions, context, and/or activities as they go about their daily lives.
- Minimize retrospective bias.
- Different kinds of questions.



https://www.researchgate.net/figure/Experience-sampling-methodology-ESM-showing-the-details-of-a-single-day-in-the-ESM_fig1_259920634

Data acquisition

- Integrated platform (app, database and reporting module)
- Uses the ESM to gain insight into people's daily life functioning.
- Enables the mapping of thoughts, feelings, experiences and behavior.
- Works on  & 
- Uses the PsyRes™ remote data infrastructure
- Developed by:
 - Department of Psychiatry and Neuropsychology of Maastricht University
 - Maastricht UMC+
 - SmarteHealth GmbH developed the PsyMate™.



Collected Data

- Three types of surveys:
 - Morning Survey (7 questions)
 - Beep Survey (38 questions)
 - Evening Survey (10 questions)
- Beep surveys:
 - User is notified 10 times a day -> fill out a short questionnaire.
 - Beeps are unevenly (randomly) spread throughout the day, though on average 90 minutes apart.



Collected Data

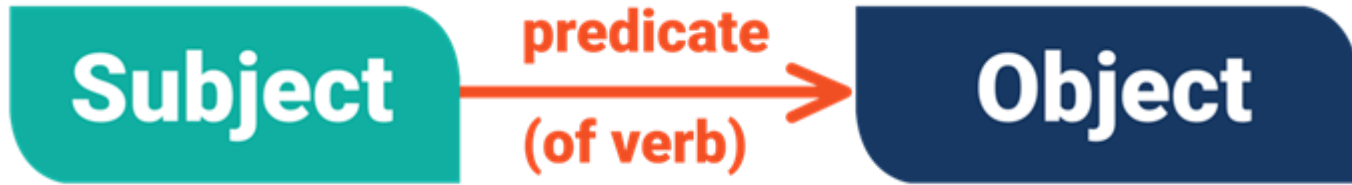
- PsyMate -> JSON format
- Collected metadata:
 - Subject ID
 - Date/Time
 - Beep number
- Collected data (questions topics):
 - Mood
 - Physical state
 - Actions
 - Thoughts
 - Location
 - Social state



Limitation of current formats

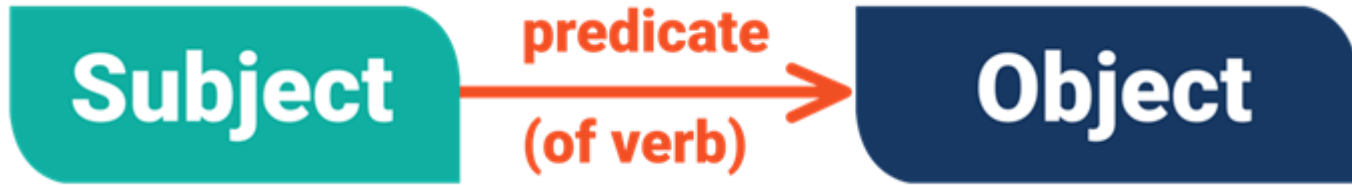
- Not machine and human interpretable (reusable?)
- Heterogeneous in nature (interoperable?)
- Various time intervals are difficult to analyze
- Lack semantics and interlinking
- FAIR data

Data representation (RDF)



- The Resource Description Framework (RDF) is a formal knowledge representation language capable of expressing a statement in a form of a triple.
- It is mature, robust, and widely used.
- Representing data in RDF, in an integrated way, allows information to be identified, disambiguated, and interconnected by software agents to read, analyze and act upon.

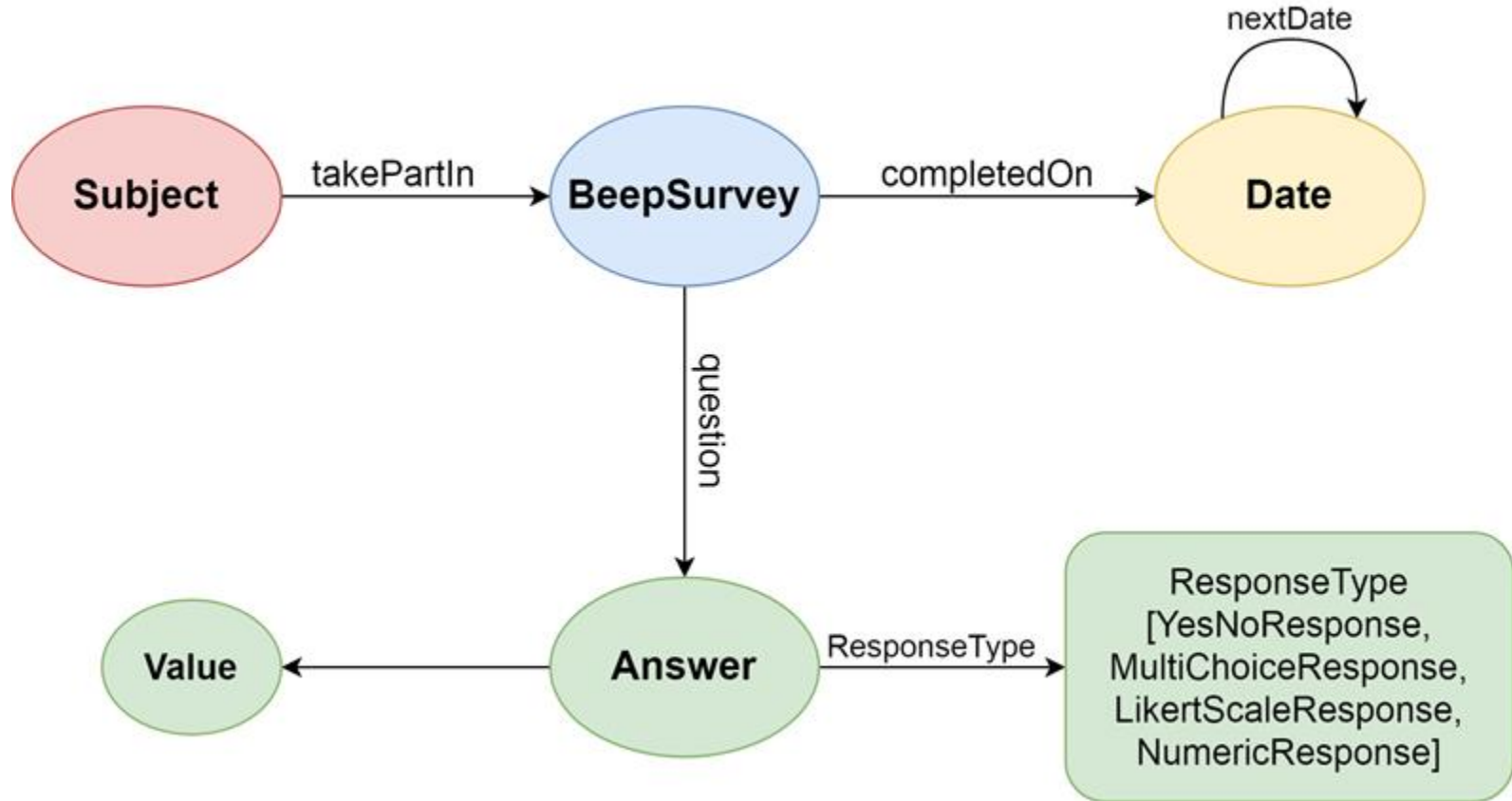
Data representation (RDF)



`<http://unimaas.nl/esm_vocabulary:survey_001>`

- Use Internationalized Resource Identifiers (IRIs; unicode) or Uniform Resource Identifiers (URIs; ascii) for names for things.
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information about them using the standards (RDF).
- Include links to other URIs so that they can discover more things.

RDF Data Model



Convert to RDF

Many tools/approaches are available:

- **General converters (multi format support)**

- RML mapper (RDF Mapping language): <https://github.com/RMLio/rmlmapper-java>
- Data2Services pipeline <https://github.com/MaastrichtU-IDS/data2services-pipeline>

- **Format specific converters**

- <https://www.w3.org/wiki/ConverterToRdf>

Query/Visualize RDF data

- RDF data is uploaded to a triple store.

- Many available solutions:

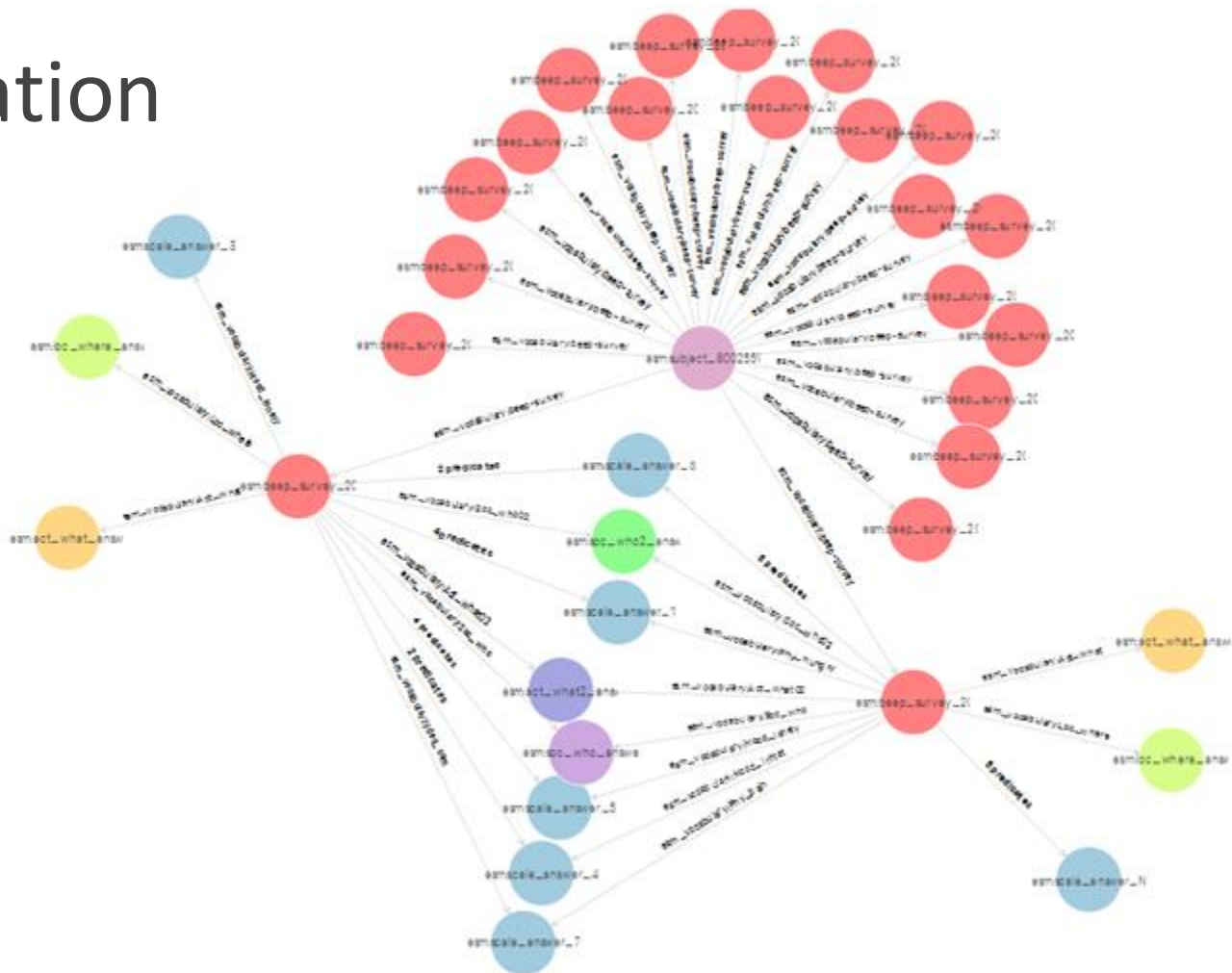


- We used GraphDB as a triple store for our dataset.

- RDF graphs are queried using SPARQL language.

- SPARQL: a semantic query language for databases—able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

Visualization



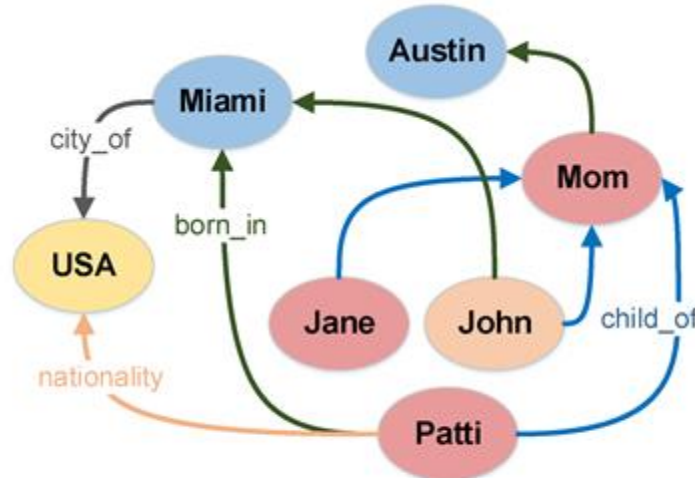
Query example

List the subjects who reported an anxiety level (scale 5 for example) and a hunger feeling (3 for example) for three consecutive days in Oct 2019.

```
1 PREFIX esm: <http://unimaas.nl/esm:>
2 PREFIX esm_vocab: <http://unimaas.nl/esm_vocabulary:>
3
4 select * where {
5
6     ?s a esm_vocab:Subject.
7
8     ?s esm_vocab:beep-survey ?p1 .
9     ?s esm_vocab:beep-survey ?p2 .
10    ?s esm_vocab:beep-survey ?p3 .
11
12    ?p1 esm_vocab:next-day ?p2 .
13    ?p2 esm_vocab:next-day ?p3 .
14
15    ?p1 esm_vocab:Mood_anxiety esm:scale_answer_5 .
16    ?p2 esm_vocab:Mood_anxiety esm:scale_answer_5 .
17    ?p3 esm_vocab:Mood_anxiety esm:scale_answer_5 .
18
19    ?p1 esm_vocab:Phy_hungry esm:scale_answer_3 .
20    ?p2 esm_vocab:Phy_hungry esm:scale_answer_3 .
21    ?p3 esm_vocab:Phy_hungry esm:scale_answer_3 .
22
23 } limit 10
```

Advanced Use Case (Rule Mining)

- AMIE+ tool, a method for rule mining from knowledge graphs, to mine patterns in the graphs.
- The tool can extract rules to be used to discriminate faithfully filled surveys from fake ones.



Logical rules

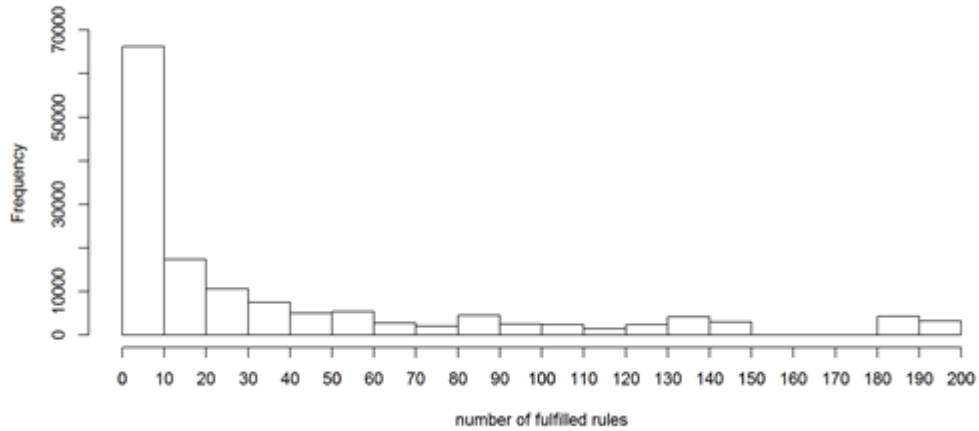
r1: $\text{born_in}(x, y) \wedge \text{city_of}(y, z) \Rightarrow \text{nationality}(x, z)$

r2: $\text{child_of}(x, y) \wedge \text{child_of}(z, y) \wedge \text{born_in}(z, p) \Rightarrow \text{born_in}(x, p)$

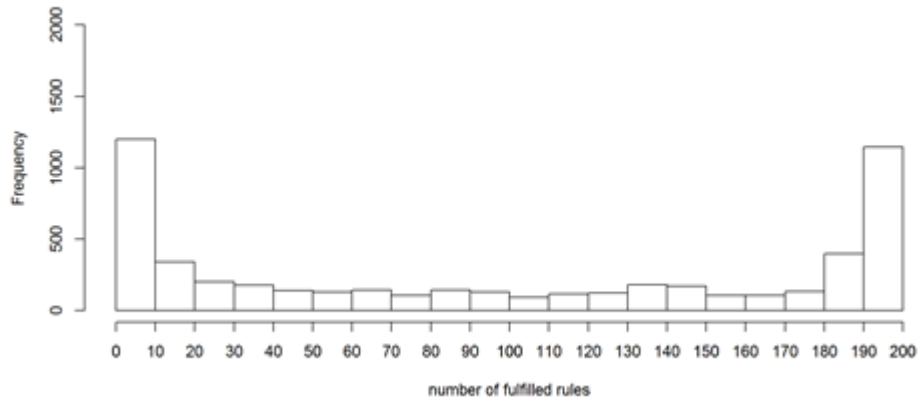
Advanced Use Case (Rule Mining)

- The assumption is that if we do not observe enough rules (patterns) in a subject/survey, it must be fake or inconsistent.
- Next, we manually checked the extracted rules whether they are meaningful to detect fake surveys and selected 200 rules to be used for detection of “scam” subjects or surveys.
- Examples:
?a <Mood_anxiety> ?b & ?a <Mood_lonely> ?b => ?a <Mood_insecur> ?b
?a <Mood_down> ?b & ?a <Phy_hungry> ?b => ?a <Mood_guilty> ?b
?a <Mood_anxiety> ?b & ?a <Phy_pain> ?b => ?a <Mood_guilty> ?b

Histogram of rules fulfilled by surveys



Histogram of rules fulfilled by subjects



Conclusion

- Semantic web technologies are valuable tools to represent ESM data
- It allows querying complex use cases and visualize the data
- Related technologies (e.g. rule mining) can be used to check/harvest “reliable data” in complex datasets (multi-protocol, multi-subject, multi-time points)